applicant on file in the Office, reference may be made to the other application and computer readable form in lieu of filing a duplicate computer readable form in the new application. The new application shall be accompanied by a letter making such reference to the other application and computer readable form, both of which shall be completely identified.

(f) In addition to the paper copy required by paragraph (c) of this section and the computer readable form required by paragraph (e) of this section, a statement that the content of the paper and computer readable copies are the same must be submitted with the computer readable form. Such a statement must be a verified statement if made by a person not registered to practice before the Office.

(g) If any of the requirements of paragraphs (b) through (f) of this section are not satisfied at the time of filing under 35 U.S.C. 111 or at the time of entering the national stage under 35 U.S.C. 371, applicant has one month from the date of a notice which will be sent requiring compliance with the requirements in order to prevent abandonment of the application. Any submission in response to a requirement under this paragraph must be accompanied by a statement that the submission includes no new matter. Such a statement must be a verified statement if made by a person not registered to practice before the Office.

(h) If any of the requirements of paragraphs (b) through (f) of this section are not satisfied at the time of filing, in the United States Receiving Office, an international application under the Patent Cooperation Treaty (PCT), applicant has one month from the date of a notice which will be sent requiring compliance with the requirements, or such other time as may be set by the Commissioner, in which to comply. Any submission in response to a requirement under this paragraph must be accompanied by a statement that the submission does not include new matter or go beyond the disclosure in the international application as filed. Such a statement must be a verified statement if made by a person not registered to practice before the Office. If applicant fails to timely provide the required computer readable form, the United States International Searching Authority shall search only to the extent that a meaningful search can be performed.

(i) Neither the presence nor the absence of information which is not required under §§ 1.821 through 1.825, in an application shall create any presumption that such information is necessary to satisfy one or more of the requirements of 35 U.S.C. 112. Further, the grant of a patent on an application that is subject to the requirements of §§ 1.821 through 1.825 shall constitute a conclusive presumption that said patent complies with the requirements of §§ 1.821 through 1.825.

(j) Envelopes containing only application papers, computer readable forms and fees filed under this section should be marked ''Box SEQUENCE.''

[55 FR 18245, May 1, 1990, as amended at 58 FR 4348, Jan. 14, 1993]

§ 1.822 Symbols and format to be used for nucleotide and/or amino acid sequence data.

(a) The symbols and format to be used for nucleotide and/or amino acid sequence data shall conform to the requirements of paragraphs (b) through (p) of this section.

(b) The code for representing the nucleotide and/or amino acid sequence characters shall conform to the code set forth in the tables in paragraphs (b)(1) and (b)(2) of this section. No code other than that specified in this section shall be used in nucleotide and amino acid sequences. A modified base or amino acid may be presented in a given sequence as the corresponding unmodified base or amino acid if the modified base or amino acid is one of those listed in paragraphs (p)(1) or (p)(2) of this section and the modification is also set forth elsewhere in the Sequence Listing (for example, FEATURES §1.823(b)(2)(ix)). Otherwise, all bases or amino acids not appearing in paragraphs (b)(1) or (b)(2) of this section shall be listed in a given sequence as ''N'' or ''Xaa,'' respectively, with further information, as appropriate, given elsewhere in the Sequence Listing.

(1) Base codes:

Symbol	Meaning
A	A; adenine
C	C; cytosine
G	G; guanine
T	T; thymine
U	U; uracil
M	A or C
R	A or G
W	A or T/U
S	C or G
Y	C or T/U
K	G or T/U
V	A or C or G; not T/U
H	A or C or T/U; not G
D	A or G or T/U; not C
B	C or G or T/U; not A
N	(A or C or G or T/U) or (unknown or other)

(2) Amino acid three-letter abbreviations:

Abbreviation	Amino acid name
Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic Acid
Asx	Aspartic Acid or Asparagine
Cys	Cysteine
Glu	Glutamic Acid
Gln	Glutamine
Glx	Glutamine or Glutamic Acid
Gly	Glycine
His	Histidine
Ile	Isoleucine
Leu	Leucine
Lys	Lysine
Met	Methionine
Phe	Phenylalanine
Pro	Proline
Ser	Serine
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine
Xaa	Unknown or other

(c) A nucleotide sequence shall be listed using the one-letter code for the nucleotide bases, as in paragraph (b)(1) of this section.

(d) The amino acids corresponding to the codons in the coding parts of a nucleotide sequence shall be typed immediately below the corresponding codons. Where a codon spans an intron, the amino acid symbol shall be typed below the portion of the codon containing two nucleotides.

(e) The amino acids in a protein or peptide sequence shall be listed using the three-letter abbreviation with the first letter as an upper case character, as in paragraph (b)(2) of this section.

(f) The bases in a nucleotide sequence (including introns) shall be listed in groups of 10 bases except in the coding parts of a sequence. Leftover bases, fewer than 10 in number, at the end of noncoding parts of a sequence shall be grouped together and separated from adjacent groups of 10 or 3 bases by a space.

(g) The bases in the coding parts of a nucleotide sequence shall be listed as triplets (codons).

(h) A protein or peptide sequence shall be listed with a maximum of 16 amino acids per line, with a space provided between each amino acid.

(i) A nucleotide sequence shall be listed with a maximum of 16 codons or 60 bases per line, with a space provided between each codon or group of 10 bases.

(j) A nucleotide sequence shall be presented, only by a single strand, in the 5′ to 3′ direction, from left to right.

(k) An amino acid sequence shall be presented in the amino to carboxy direction, from left to right, and the amino and carboxy groups shall not be presented in the sequence.

(l) The enumeration of nucleotide bases shall start at the first base of the sequence with number 1. The enumeration shall be continuous through the whole sequence in the direction 5′ to 3′. The enumeration shall be marked in the right margin, next to the line containing the one-letter codes for the bases, and giving the number of the last base of that line.

(m) The enumeration of amino acids may start at the first amino acid of the first mature protein, with number 1. The amino acids preceding the mature protein, e.g., pre-sequences, pro-sequences, pre-pro-sequences and signal sequences, when presented, shall have negative numbers, counting backwards starting with the amino acid next to number 1. Otherwise, the enumeration of amino acids shall start at the first amino acid at the amino terminal as number 1. It shall be marked below the sequence every 5 amino acids.

(n) For those nucleotide sequences that are circular in configuration, the enumeration method set forth in paragraph (l) of this section remains applicable with the exception that the designation of the first base of the nucleotide sequence may be made at the option of the applicant. The enumeration method for amino acid sequences that is set forth in paragraph

(m) of this section remains applicable for amino acid sequences that are circular in configuration.

(o) A sequence with a gap or gaps shall be presented as a plurality of separate sequences, with separate sequence identifiers, with the number of separate sequences being equal in number to the number of continuous strings of sequence data. A sequence that is made up of one or more noncontiguous segments of a larger sequence or segments from different sequences shall be presented as a separate sequence.

(p) The code for representing modified nucleotide bases and modified and unusual amino acids shall conform to the code set forth in the tables in paragraphs (p)(1) and (p)(2) of this section. The modified base controlled vocabulary in paragraph (p)(1) of this section and the modified and unusual amino acids in paragraph (p)(2) of this section shall not be used in the nucleotide and/or amino acid sequences; but may be used in the description and/or the ''Sequence Listing'' corresponding to, but not including, the nucleotide and/or amino acid sequence.

(1) Modified base controlled vocabulary:

Abbreviation	Modified base description
ac4c	4-acetylcytidine.
chm5u	5-(carboxyhydroxylmethyl)uridine.
cm	2′-O-methylcytidine.
cmnm5s2u	5-carboxymethylaminomethyl-2-thioridine.
cmnm5u	5-carboxymethylaminomethyluridine.
d	dihydrouridine.
fm	2′-O-methylpseudouridine.
galq	beta,D-galactosylqueosine
gm	2′-O-methylguanosine.
i	inosine.
i6a	N6-isopentenyladenosine.
m1a	1-methyladenosine.
m1f	1-methylpseudouridine.
m1g	1-methylguanosine.
ml1	1-methylinosine.
m22g	2,2-dimethylguanosine.
m2a	2-methyladenosine.
m2g	2-methylguanosine.
m3c	3-methylcytidine.
m5c	5-methylcytidine.
m6a	N6-methyladenosine.
m7g	7-methylguanosine.
mam5u	5-methylaminomethyluridine.
mam5s2u	5-methoxyaminomethyl-2-thiouridine.
manq	beta,D-mannosylqueosine.
mcm5s2u	5-methoxycarbonylmethyluridine.
mo5u	5-methoxyuridine.
ms2i6a	2-methylthio-N6-isopentenyladenosine.
ms2t6a	N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine.
mt6a	N-((9-beta-D-ribofuranosylpurine-6-yl)N-methyl-carbamoyl)threonine.
mv	uridine-5-oxyacetic acid methylester.
o5u	uridine-5-oxyacetic acid (v).
osyw	wybutoxosine.
p	pseudouridine.
q	queosine.
s2c	2-thiocytidine.
s2t	5-methyl-2-thiouridine.
s2u	2-thiouridine.
s4u	4-thiouridine.
t	5-methyluridine.
t6a	N-((9-beta-D-ribofuranosylpurine-6-yl)carbamoyl)threonine.
tm	2′-O-methyl-5-methyluridine.
um	2′-O-methyluridine.
yw	wybutosine.
x	3-(3-amino-3-carboxypropyl)uridine, (acp3)u.

(2) Modified and unusual amino acids:

Abbreviation	Modified and unusual amino acid
Aad	2-Aminoadipic acid.
bAad	3-Aminoadipic acid.
bAla	beta-Alanine,beta-Aminopropionic acid.
Abu	2-Aminobutyric acid.
4Abu	4-Aminobutyric acid, piperidinic acid.
Acp	6-Aminocaproic acid.
Ahe	2-Aminoheptanoic acid.
Aib	2-Aminoisobutyric acid.
bAib	3-Aminoisobutyric acid.
Apm	2-Aminopimelic acid.
Dbu	2,4-Diaminobutyric acid.
Des	Desmosine.
Dpm	2,2′-Diaminopimelic acid.
Dpr	2,3-Diaminopropionic acid.
EtGly	N-Ethylglycine.
EtAsn	N-Ethylasparagine.
Hyl	Hydroxylysine.
aHyl	allo-Hydroxylysine.
3Hyp	3-Hydroxyproline.
4Hyp	4-Hydroxyproline.
Ide	Isodesmosine.
alle	allo-Isoleucine.
MeGly	N-Methylglycine, sarcosine.
Melle	N-Methylisoleucine.
MeLys	N-Methylvaline.
Nva	Norvaline.
Nle	Norleucine.
Orn	Ornithine.

§ 1.823 Requirements for nucleotide and/or amino acid sequences as part of the application papers.

(a) The ''Sequence Listing,'' required by § 1.821(c), setting forth the nucleotide and/or amino acid sequences, and associated information in accordance with paragraph (b) of this section, must begin on a new page and be titled ''Sequence Listing'' and appear immediately prior to the claims. Each page of the ''Sequence Listing'' shall contain no more than 66 lines and each line shall contain no more than 72 characters. A fixed-width font shall be